Journal of High Speed Networks, 5(4):347--363, 1996

Analysis of One-Way Reservation Algorithms

Israel Cidon
Sun Microsystems
Mountain View, CA 94043-1100
and
Department of Electrical Engineering
Technion, Haifa, Israel

*Raphael Rom**
Department of Electrical Engineering
Technion, Haifa, Israel
and
Sun Microsystems
Mountain View, CA 94043-1100

Yuval Shavitt
Department of Electrical Engineering
Technion, Haifa, Israel

January 6, 1995

Abstract

Modern high speed networks (and ATM networks in particular) use resource reservation as a basic means of establishing communications. One-way on-the-fly reservation is a method for allocating resources for short bursts of data when regular reservation is too costly. The first such algorithms were recently suggested by Turner. We investigate two examples that are characteristic to the way traffic streams interact in virtual circuit networks: a three node subnetwork that also acts as a 4X2 switch and a ring. For both systems we compute system throughput under homogeneous load, and compare it with the throughput when fast reservation protocols (FRP) are used. For the three node subnetwork we give an explicit expression for an upper bound.

*This work was supported in part by the Technion V.P.R. fund and the Fund for the Promotion of Research at the Technion.

Journal of High Speed Networks, 5(4):347--363, 1996

1 Introduction

High speed networks are intended to support applications with widely varying traffic characteristics: from short database queries to long video streams. In order to use the network resources efficiently, bandwidth reservations are made to ensure high probability of data arrival to their destinations. For applications such as constant bit rate video or voice conversations this is the right approach. However, for bursty traffic, i.e., traffic whose intensity varies in time, reservation itself introduces non-negligible overhead.

Short bursts are those whose transmission time is not more than a few round trip delays. For such bursts waiting for a reservation, that itself takes a few round trip delays, is clearly not acceptable. The best end-to-end method for this type of bursts is to send the data without reservation and use time-outs (possibly at a higher layer) to detect failures. However, if several bursts arrive to an ATM switch at the same time and compete for buffers they may disturb each other. To avoid this, Turner [Tur92] suggested to use one-way on-the-fly reservation. In his scheme, a burst that arrives to an ATM switch and finds a sufficient number of buffers reserves the required buffer space to prevent disturbance from new coming bursts. This process is repeated in every switch along the burst route. Note that the scheme does not guarantee that a burst that succeeds in reserving enough bandwidth in one switch will also succeed in the next one along the route. There is also no mechanism to release immediately the partial bandwidth already reserved (see below). The same solution does not fit longer bursts. Here, the overhead of reservation is not as bothersome so two-way reservation algorithms can be used. Widjaja [Wid94] computed the break-even point where the performance of the two schemes is equal.

In this work we give an exact performance analysis of one-way on-the-fly reservation algorithms and compare the results with the performance of fast-reservation protocols (FRP). Our analysis is based on modeling bursts and bandwidth on a link as customers and G/M/N/N queues, respectively [Wid94, HLNP93, CLG94]. We are able, then, to use results that were obtained from the analysis of telephone switching networks [Gir90]. However, note that bursts that use one-way reservation do not behave as telephone calls. Specifically, if a burst is rejected by a switch in the middle of its route the burst holds all the switches along the route, from the source to the blocking switch, for the entire burst duration, while in telephone switching networks a failure to capture a server in one of the switches along a call route causes the immediate release of those servers that the call managed to capture before being blocked. In the model considered by many to analyze circuit switched networks in general and telephone switching networks in particular (e.g., [**?**, **?**]), one cannot introduce dependency between the system state (the number of busy servers in each link) and the group of links an incoming call captures. In these models, the group ofcaptured links is always predefined. Consequently, much of the research done on circuit-switched networks [**?**, **?**] cannot be applied to the analysis of one-way on-the-fly reservation algorithms.

The main difficulty in using results from the telephony model is that the process that describes the traffic out of a telephone switch (a link in our case), the *carried traffic* in telephony jargon, is not a renewal process. Widjaja [Wid94] bypassed this difficulty by assuming that the rejection probabilities in the links along the burst route are independent. We show in section 3 that this assumption is incorrect, and demonstrate the inaccuracies that result from such an assumption. In our analysis, we look at Markov chains that represent the state of the full system and thus avoid the need to explicitly describe the traffic out of every switch. Recently, Cho and Leon-Garcia [CLG94] analyzed a system with only one potential congested link. Doing so lead them to erroneously state that one-way reservation and fast reservation perform equally when the propagation delay is zero. This statement is incorrect when more then a single link can potentially be congested. Thus, their simplifying assumption caused them to miss an important deficiency of on-the-fly reservation.

Two special cases, a three-node subnetwork and a ring of nodes, are analyzed. These are two typical examples of interaction among bursts that travel more than a single hop. Our analysis shows that when one-way reservation is used, the inability to throttle a burst source that fails to capture bandwidth along its entire route significantly damages the throughput. In the ring example, the throughput loss is more apparent, but the upper bound calculation for the three-node subnetwork demonstrates that this throughput loss is intrinsic even for a simple interaction of streams. In scenarios where the travel hop distance grows, simulations show even a greater loss of throughput. In Section 4 we compare the performance between fast reservation and one-way reservation. Note that for fast-reservation protocols the work on circuit-switched networks can be applied.

The Model

The following model applies to the two analyzed systems. The input traffic to the systems is comprised of input streams. Each stream is comprised of constant bit rate bursts whose duration is exponentially distributed with mean $1/\mu$. The interarrival time between the bursts of stream i is exponentially distributed with mean $1/\lambda_i$. The links in the system can each support up to N bursts simultaneously, and are modeled as N server queues with no waiting room (G/M/N/N queue). The bursts are modeled as customers that try to capture a server (bandwidth) for the duration of the burst. Bursts that arrive to a link and find no available bandwidth are lost. Note that the use of queueing theory terms might be misleading since our queues do not form a queueing system per se. A burst (customer) that arrives to the system tries to capture bandwidth simultaneously for each link on its route. Only bursts that succeed in the entire reservation contribute to the useful throughput.

The systems we analyze have a finite (and small) number of links. The links are modeled by G/M/N/N queues and thus the total number of system states is finite. In addition, the periods until a state transition occurs are all exponentially distributed. As a result, we can model the system by finite continuous time Markov chains, each is uniquely described by the rate transition matrix, Q. The element $q_{i,j}$ in Q is the transition rate from state i to state j, the negation of a diagonal element, $-q_{i,i}$, is the output rate from state i.

2 Analysis of a Three Node Subnetwork

2.1 Exact Analysis

We examine a three node subnetwork (Figure 1) with four external input streams. Cells of streams 1 and 3 are routed to link 1, while cells of streams 2 and 4 are routed to link 2. The subnetwork can be viewed as a large switch with 4 inputs and 2 outputs, a typical scenario for connection oriented network.

The interarrival periods between bursts in the four input streams are exponentially distributed with mean $1/\lambda_i$, $1 \leq i \leq 4$. We define this system state by the six-tuple comprised of the following variables:

- The number of bursts that are transmitted over link 3.

Journal of High Speed Networks, 5(4):347--363, 1996

- The number of bursts that are transmitted over link 4.
- The number of bursts that are transmitted over links 3 and 1.
- The number of bursts that are transmitted over links 4 and 1.
- The number of bursts that are transmitted over links 3 and 2.
- The number of bursts that are transmitted over links 4 and 2.

Since the number of bursts each link carries is between 0 and N, the number of system states is $(N+1)^6$. In addition, since the transitions among states is memoryless the system is Markovian. The large number of states makes the analytic or numeric solution of this system laborious even for a small N. However, when the system is symmetric, i.e., $\lambda_i = \frac{\lambda}{2}$ for all i, we can disregard link 2, use a four-tuplet state descriptor, and reduce the number of states to $(N+1)^4$.

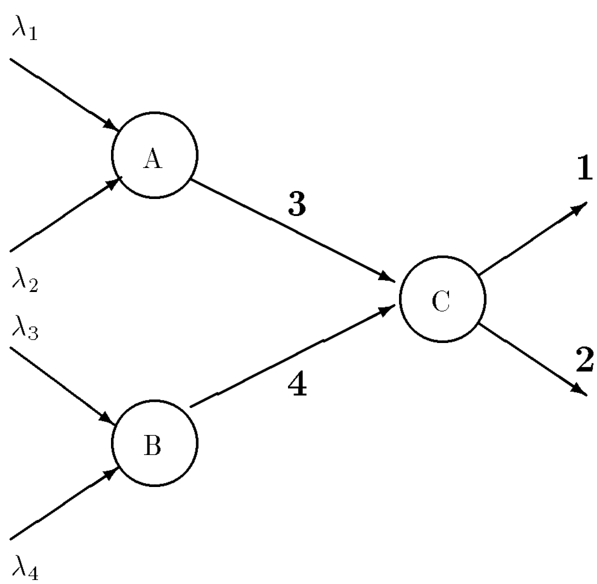

Figure 1: A three node subnetwork

We wish to calculate the system utilization as a function of the load, $\rho = \lambda/(N\mu)$. The system utility is defined as

$$U^* = \lim_{T\to\infty} \frac{1}{2NT} \int_0^T (n_{c_1}(t) + n_{c_2}(t))dt = \lim_{T\to\infty} \frac{1}{NT} \int_0^T n_{c_1}(t)dt \tag{1}$$

where $n_{c_1}(t)$ and $n_{c_2}(t)$ are the number of bursts that are transmitted over links 1 and 2 at time t, respectively. Since the Markov chain that models the system is finite with exponential transitions, it is ergodic, meaning that we can replace the average utility in time with the utility probability at every moment. In our case we can replace the calculation in equation 1 with the expected number of bursts in link 1.

Consider now a state $\langle i,j,l,m\rangle$ where $i < N$ and $j < N$ bursts are occupying links 3 and 4, respectively, and $l \le i$ and $m \le j$ out of them are occupying link 1, as well. Several independent transitions may occur: a burst may leave links 3 and 1, a burst may leave links 4 and 1, a burst may leave link 3 and not leave link 1, a burst may leave link 4 and not leave link 2, a burst may arrive to links 3 and 1, a burst may arrive to links 3 and 2, a burst may arrive to links 4 and 1, and a burst may arrive to links 4 and 2. All the transitions are exponentially distributed with rates $l\mu$, $m\mu$, $(i-l)\mu$, $(j-m)\mu$, $\lambda/2$, $\lambda/2$, $\lambda/2$, and $\lambda/2$, correspondingly. We can write the elements of the transition rate matrix, Q:

$$q_{\langle i,j,l,m\rangle,\langle i-1,j,l-1,m\rangle} = l\mu \tag{2}$$
$$q_{\langle i,j,l,m\rangle,\langle i,j-1,l,m-1\rangle} = m\mu \tag{3}$$
$$q_{\langle i,j,l,m\rangle,\langle i-1,j,l,m\rangle} = (i-l)\mu \tag{4}$$
$$q_{\langle i,j,l,m\rangle,\langle i,j-1,l,m\rangle} = (j-m)\mu \tag{5}$$

For the case where $l+m < N$:

$$q_{\langle i,j,l,m\rangle,\langle i+1,j,l+1,m\rangle} = \lambda/2 \tag{6}$$
$$q_{\langle i,j,l,m\rangle,\langle i+1,j,l,m\rangle} = \lambda/2 \tag{7}$$
$$q_{\langle i,j,l,m\rangle,\langle i,j+1,l,m+1\rangle} = \lambda/2 \tag{8}$$
$$q_{\langle i,j,l,m\rangle,\langle i,j+1,l,m\rangle} = \lambda/2 \tag{9}$$

For the case where $l+m = N$:

$$q_{\langle i,j,l,m\rangle,\langle i+1,j,l,m\rangle} = \lambda \tag{10}$$
$$q_{\langle i,j,l,m\rangle,\langle i,j+1,l,m\rangle} = \lambda \tag{11}$$

When i, j, or both are equal to N some of the transitions in equations 6 - 11 become void since they lead to nonexisting states. The diagonal of Q holds, as defined above, the rate to move out of states. We can write then:

$$q_{\langle i,j,l,m\rangle,\langle i,j,l,m\rangle} = -[2\lambda + (i+j)\mu] \tag{12}$$

For the case where $i = N$ or $j = N$ but not both:

$$q_{\langle i,j,l,m\rangle,\langle i,j,l,m\rangle} = -[\lambda + (i+j)\mu] \tag{13}$$

For the case where $i = j = N$:

$$q_{\langle N,N,l,m\rangle,\langle N,N,l,m\rangle} = -2N\mu \tag{14}$$

We can now calculate the steady state probabilities $\pi_{\langle i,j,l,m\rangle}$ using the relation:

$$\vec{\pi}Q = 0 \tag{15}$$

and the probability conservation relation:

$$\sum_{\substack{i,j,l,m\in[0,\ 1,\ \ldots,\ N]\\ l\le i,\ m\le j,\ l+m\le N}} \pi_{\langle i,j,l,m\rangle} = 1 \tag{16}$$

where $\vec{\pi}$ is the vector of steady state probabilities.

The system utilization U^* is given by

$$U^* = \frac{1}{N} \sum_{\substack{i,j,l,m \in [0,\ 1,\ \ldots,\ N] \\ l \le i,\ m \le j,\ l+m \le N}} (l+m)\pi_{\langle i,j,l,m \rangle} \tag{17}$$

Figure 2 shows the system utilization as a function of the load, ρ.

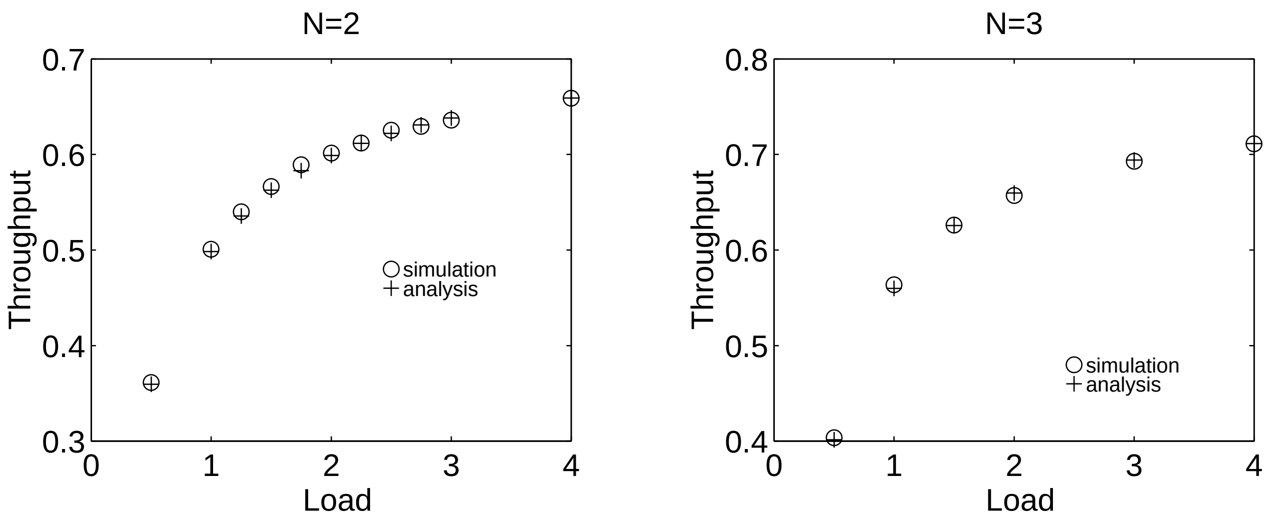

Figure 2: The three node subnetwork utilization (U) as a function of the load (ρ).

2.2 Approximated Analysis

The above exact analysis requires the solution of a set of $O((N+1)^4)$ equations which becomes burdensome for large N. To make the analysis more tractable, we assume that the burst duration at the two links along its path is independent. This is typical to circuit-switching network analysis and simulations show that it does not impact the final results for large N values[1]. We would like to write an expression for the input process to link 1 that is based on the states of links 3 and 4 and then to analyze the G/M/N/N queue that represents link 1. However, as we already stated, the input to link 1 is not a renewal process since the interarrival period distribution depends on the states of queues 3 and 4. Nevertheless, we assume that the process is a renewal process with interarrival periods that are distributed according to a weighted sum of the distributions of the interarrival periods in all the states of queues 3 and 4 (see equation 19). Our assumption can be easily justified for extreme load situations: when the system is lightly loaded the links are almost never full and the interarrival time is exponentially distributed with mean $1/\lambda$; when the system is heavily loaded the queues are almost always full and the interarrival periods are determined by the server release process and are therefore exponentially distributed with mean $1/(N\mu)$. Simulations justify our assumption.

Using our independence and renewal assumptions, links 3 and 4 can be modeled as M/M/N/N queues with a Poisson arrival process with rates $\lambda_1 + \lambda_2 = \lambda$ and $\lambda_3 + \lambda_4 = \lambda$, respectively, and link

[1]The assumption is similar in spirit to Kleinrock's independence assumption [Kle76] for packet-switching networks.

1 can be modeled as a G/M/N/N queue whose arrival process depends on the states of links 3 and 4. The steady state probability to find links 3 and 4 in state (i, j) is $p_i p_j$, where p_k is the probability of an M/M/N/N queue with N exponential servers with parameter μ and a Poisson input stream with parameter λ to have k busy servers [Kle75, ch. 3]:

$$p_k = \frac{\frac{(\lambda/\mu)^k}{k!}}{\sum_{i=0}^{N} \frac{(\lambda/\mu)^i}{i!}} \tag{18}$$

Using the steady state probabilities we write $\phi(s)$, the Laplace transform of the probability density function (pdf) of the interarrival times of bursts to link 1, as a function of $\phi_{i,j}(s)$, the Laplace transform of the pdf of the interarrival times of bursts to link 1 given that i and j bursts are already occupying links 3 and 4, respectively. Note that due to system symmetry, $\phi_{i,j}(s) = \phi_{j,i}(s)$ for all i and j. We clearly have

$$\phi(s) = \sum_{i=0}^{N} p_i \sum_{j=0}^{N} p_j \phi_{i,j}(s) \tag{19}$$

Consider now the situation where $i < N$ and $j < N$ bursts are already occupying links 3 and 4. The following independent events may occur, each after an exponentially distributed period:

- a burst may leave link 3, with rate $i\mu$
- a burst may leave link 4, with rate $j\mu$
- a burst may arrive to link 3 from input process 1, with rate $\lambda/2$
- a burst may arrive to link 3 from input process 2, with rate $\lambda/2$
- a burst may arrive to link 4 from input process 3, with rate $\lambda/2$
- a burst may arrive to link 4 from input process 4. with rate $\lambda/2$

The interval to the next event is therefore also exponentially distributed and has a Laplace transform given by $(2\lambda + (i+j)\mu)/(2\lambda + (i+j)\mu + s)$. Thus, for $0 \le i < N$ and $0 \le j < N$ we have:

$$\begin{aligned} \phi_{i,j}(s) &= \frac{\lambda}{2\lambda + (i+j)\mu + s} + \frac{\lambda/2}{2\lambda + (i+j)\mu + s}\phi_{i+1,j}(s) + \frac{\lambda/2}{2\lambda + (i+j)\mu + s}\phi_{i,j+1}(s) \\ &\quad + \frac{i\mu}{2\lambda + (i+j)\mu + s}\phi_{i-1,j}(s) + \frac{j\mu}{2\lambda + (i+j)\mu + s}\phi_{i,j-1}(s) \\ &= \frac{\lambda}{2\lambda + (i+j)\mu + s} + \frac{\lambda/2}{2\lambda + (i+j)\mu + s}\left[\phi_{i+1,j}(s) + \phi_{i,j+1}(s)\right] \\ &\quad + \frac{\mu}{2\lambda + (i+j)\mu + s}\left[i\phi_{i-1,j}(s) + j\phi_{i,j-1}(s)\right] \end{aligned} \tag{20}$$

In equation 20 above, each term is a multiplication of the probability that a certain event is the next and the Laplace transform of the distribution of the time to the next event. For example, the first term is a multiplication of the probability that the next event is an arrival of a burst destined to link 1, $\lambda/(2\lambda + (i+j)\mu)$, and the Laplace transform of the distribution of the time to the next event, $(2\lambda + (i+j)\mu)/(2\lambda + (i+j)\mu + s)$.

For the special boundary cases we have:

$$\phi_{N,i}(s) = \frac{\lambda/2}{\lambda+(N+i)\mu+s} + \frac{\lambda/2}{\lambda+(N+i)\mu+s}\phi_{N,i+1}(s) + \frac{\mu}{\lambda+(N+i)\mu+s}\left[N\phi_{N-1,i}(s)+i\phi_{N,i-1}(s)\right] \quad 0 \le i < N \tag{21}$$

$$\phi_{N,N}(s) = \frac{N\mu}{2N\mu+s}\left[\phi_{N,N-1}(s)+\phi_{N-1,N}(s)\right] \tag{22}$$

To express $\phi_{i,j}(s)$ as explicit functions, we must solve this system of $(N+1)^2$ linear equations with $(N+1)^2$ variables.[2] The corresponding matrix contains non-zero elements only in the three main diagonals and in two more diagonals that are at distance $N+1$ from the main diagonal. This system can be easily solved by expressing the variables from position $N+2$ and on by the first $N+1$ variables, a process whose complexity is linear by the number of variables, i.e., $O((N+1)^2)$. At the end of this process we are left with $N+1$ equation that can be solved in $O(N^{2+\alpha})$, where α is currently below 0.5. Then, the values for the this variables is substituted in the expressions for the other $(N+1)^2-(N+1)$ variables. This last process is of order $O(N^3)$ which is also the complexity order for the whole process. In the end we obtain $\phi(s)$, the Laplace transform of the pdf of the interarrival times of the input stream to the G/M/N/N server that represents link 1.

The system utilization at an arbitrary point in time is given by the utilization of the G/M/N/N queue representing link 1:

$$U^* = \sum_{i=0}^{N} \frac{i}{N} p_i^* \tag{23}$$

where p_k^* is the probability to find k busy servers in queue representing link 1 at an arbitrary point in time. Using the relation [Tak62, ch. 4]

$$p_k^* = \frac{p_{k-1}}{-k\mu\phi'(0)} \qquad 0 < k \le N \tag{24}$$

we express p_k^* as a function of the number of busy servers at the arrival epochs, and substituting this in equation 23 we obtain

$$U^* = \frac{1-B}{-N\mu\phi'(0)} \tag{25}$$

where B is the blocking probability for this queue, i.e., the probability that a burst that passes through link 3 or 4 is rejected by the link 1.

To avoid the need to derive ϕ, one can approximate U^* by U, the system utilization at the arrival epochs [Gir90, ch. 3]:

$$U = \sum_{i=0}^{N} \frac{i}{N} p_i = \frac{h_1}{N}\left(1 - \left(\sum_{j=0}^{N} \binom{N}{j} \frac{1}{h_j}\right)^{-1}\right) = \frac{1}{N}(1-B)\frac{\phi(\mu)}{1-\phi(\mu)} \tag{26}$$

where

$$B = \left[\sum_{j=0}^{N} \binom{N}{j} \frac{1}{h_j}\right]^{-1} \tag{27}$$

[2]The system size can be reduced to $(N+1)(N+2)/2$ by eliminating symmetric states

$$h_j = \prod_{i=1}^{j} \frac{\phi(i\mu)}{1-\phi(i\mu)} \quad j \geq 1 \tag{28}$$

$$h_0 = 1 \,. \tag{29}$$

For the values of N we computed the graphs for U and U^* were hardly distinguishable.

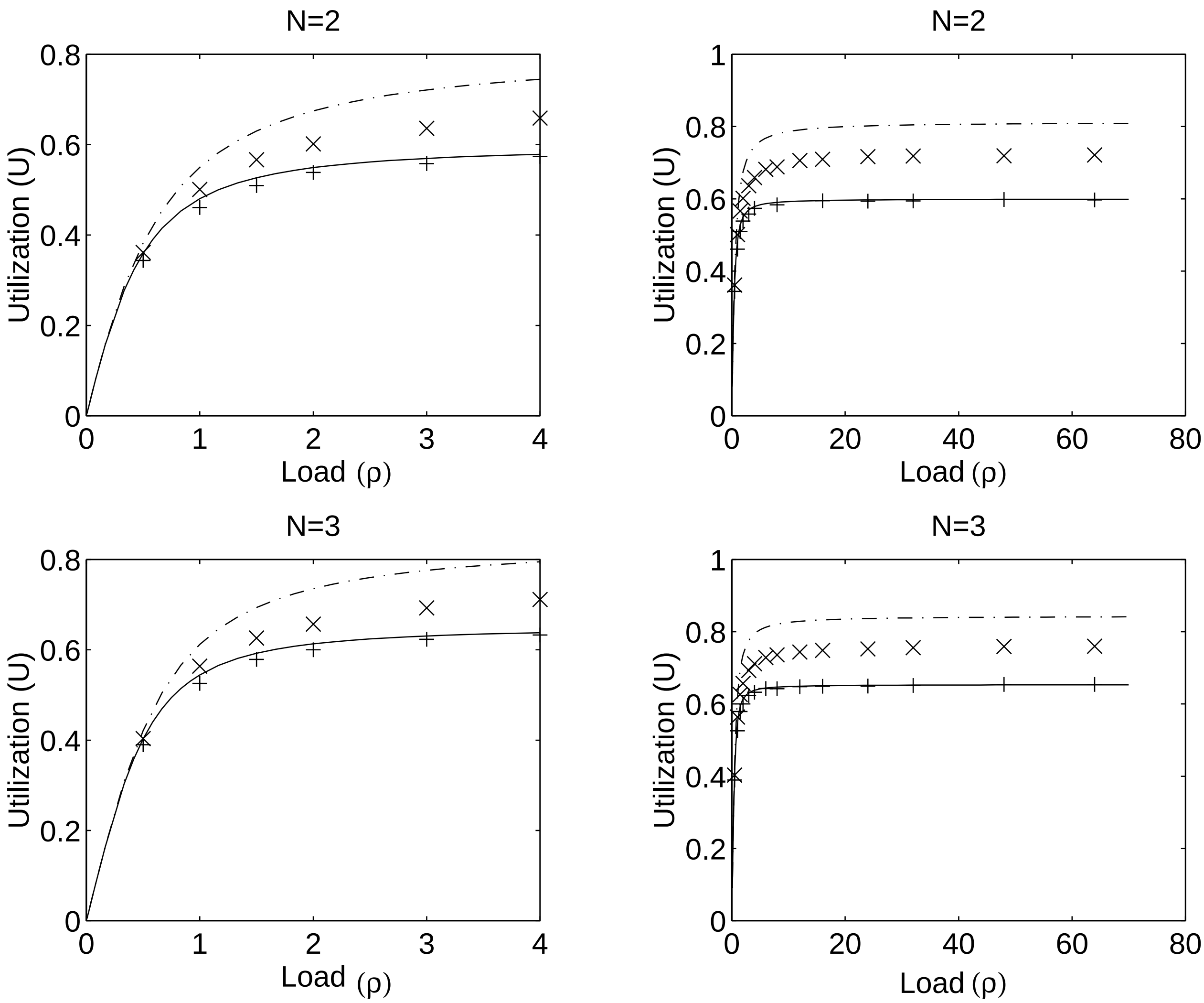

Figure 3: Comparison of simulation results with and without the independence assumption

Figure 3 shows the utilization of link 1 as a function of the load, ρ, for the cases where the number of servers in each link is two and three. The figures compare the analysis (solid line) with simulation results of two systems: (a) a system where the burst duration is exponentially distributed but without the independence assumption, i.e., a burst occupying two links occupies both the same time (the simulation points are plotted as Xs); and (b) a system where the burst duration is independently selected for each link (the simulation points are plotted as plus signs). Note that the analysis and the simulation points of the second system are very close which confirms our analysis under the independence assumption. Figure 4 compares the two simulated systems when the number of servers

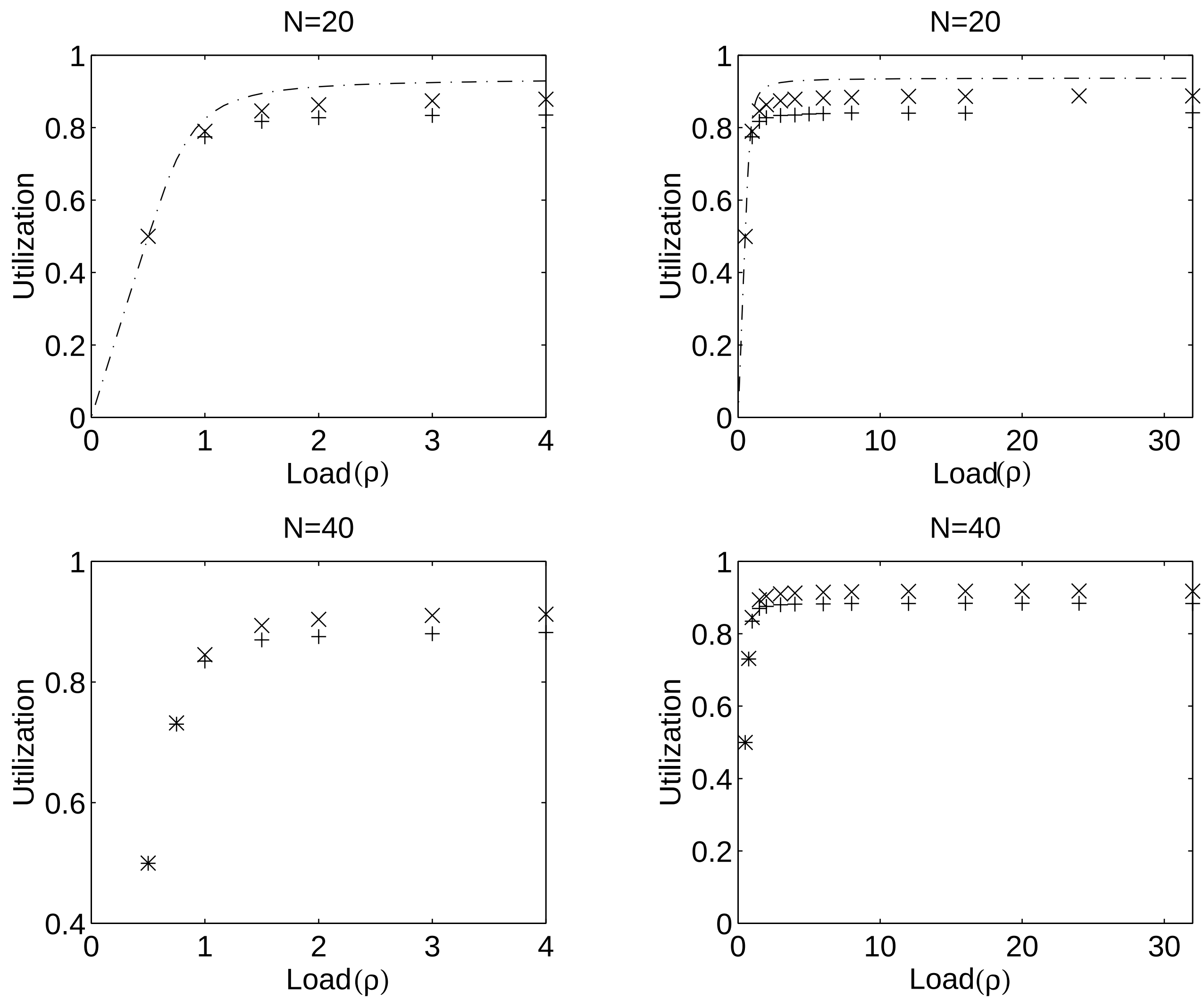

Figure 4: Comparison of simulation results with and without the independence assumption

is 20 and 40. The difference in the utility of the two systems is less than 5% for $N = 20$ and about 3% for $N = 40$, which validates the independence assumption for a large number of servers.

2.3 Upper Bounds on the Utilization

In this section we develop upper bounds on the utilization whose calculation is simpler than the one for the approximated analysis. First, we develop an upper bound for heavy loads whose calculation requires a linear number of multiplications. Then we give a tighter upper bound which is a function of the load, ρ, and that requires $O(N^4)$ multiplications.

Let $n_1(t)$, $n_2(t)$, $n_3(t)$, and $n_4(t)$ be the number of bursts from input streams 1, 2, 3, and 4, respectively, that occupy the servers of links 3 and 4 at time t; $n_1(t)+n_2(t) \leq N$ and $n_3(t)+n_4(t) \leq N$. For every time t, The number of bursts that succeed to capture bandwidth at link 1 is $n_{c_1}(t) \leq$

$\min\{n_1(t) + n_3(t),\ N\}$. From equation 1 we can write:

$$U^* = \lim_{T\to\infty} \frac{1}{NT}\int_0^T n_{c_1}(t)dt \le \lim_{T\to\infty} \frac{1}{NT}\int_0^T \min\{n_1(t)+n_3(t),\ N\}dt \triangleq \overline{U} \tag{30}$$

Since the system is ergodic, we can replace the average utility in time with the utility probability at every moment. In our case we can replace the calculation in equation 30 with the expected number of bursts in links 3 and 4 that originate from streams 1 and 3, respectively. We assume that the system is heavily loaded, i.e., whenever a server in links 3 or 4 is freed a new burst captures it again immediately. Since the arrival rates of the four input streams are equal, the probability that the evacuated stream is captured by a burst that is destined to link 1 is 1/2. The heavy load assumption allows us to assume that at all times $n_1(t) + n_2(t) = N$ and $n_3(t) + n_4(t) = N$. From the system symmetry it follows that at every time t the probability that $n_1(t) + n_3(t) = i$ and $n_2(t) + n_4(t) = 2N - i$ is:

$$\binom{2N}{i}\frac{1}{2^{2N}} \tag{31}$$

Which leads to the following expression for the expected number of bursts in link 1 that originated from streams 1 and 3:

$$\begin{aligned}
\overline{U}(N) &= \frac{1}{N2^{2N}}\sum_{i=0}^{2N}\binom{2N}{i}\min\{i, N\} \\
&= \frac{1}{N2^{2N}}\left[\sum_{i=1}^{N}\binom{2N}{i} i + \sum_{i=N+1}^{2N}\binom{2N}{i} N\right] \\
&= \frac{1}{N2^{2N}}\left[2N\sum_{i=0}^{N-1}\binom{2N-1}{i} + N\frac{1}{2}\left(2^{2N} - \binom{2N}{N}\right)\right] \\
&= 1 - \binom{2N}{N}/2^{2N+1}
\end{aligned} \tag{32}$$

Using Stirling's approximation for the factorial it can be shown that $\overline{U}(N)$ is approaching 1 as N grows in the same rate as $1 - 1/(2\sqrt{\pi N})$. For the values of N we plotted in figures 3 and 4 our upper bound obtain the values: $\overline{U}(N=2) = 0.8125$, $\overline{U}(N=3) = 0.84375$, $\overline{U}(N=20) \approx 0.937$, and $\overline{U}(N=40) \approx 0.956$.

Figure 4 demonstrates that this upper bound is reasonable only for $\rho > 10$ and is not tight enough for lower loads. To make it tighter, the load on the system should be considered, i.e., the fact that there is a probability that servers are not occupied. The probability to have i servers occupied, p_i, in links 3 and 4 is given by the Erlang B formula (equation 18). We use this probability together with the technique used in the derivation of equation 32 to obtain the following upper bound:

$$\overline{U}(N,\rho) = \frac{1}{N}\sum_{i=0}^{N}\sum_{k=0}^{N} p_0^2\left(\frac{N\rho}{2}\right)^{i+k}\frac{1}{i!k!}\sum_{j=0}^{i}\sum_{l=0}^{k}\binom{i}{j}\binom{k}{l}\min\{j+l,\ N\} \tag{33}$$

The dashed lines in figures 3 and 4, show $\overline{U}(N,\rho)$. $\overline{U}(N,\rho)$ approaches $\overline{U}(N)$ for small values of N and $\rho > 10$. For greater values of N, $\overline{U}(N,\rho)$ approaches $\overline{U}(N)$ for lower ρ values. But, for all the values of N and ρ, $\overline{U}(N,\rho) < \overline{U}(N)$.

3 Analysis of a Cycle of Streams

Another topology we examine is a ring network with n nodes, each receiving an external Poisson input stream with rate λ_0 from a source outside the ring and trying to transmit it to a node h hops away clockwise (Figure 5). We analyze the case when $h = 2$ using similar techniques to the ones used in the previous section, i.e., links are modeled as G/M/N/N queues and bursts that arrive to a link are modeled as customers that try to capture a server.

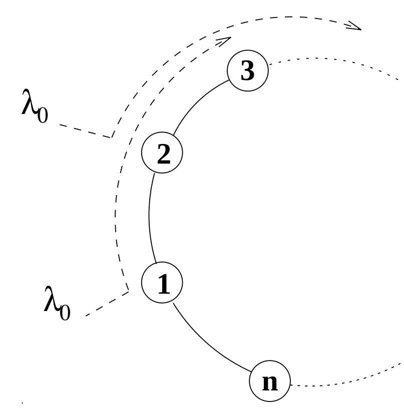

Figure 5: A ring with two-hop streams

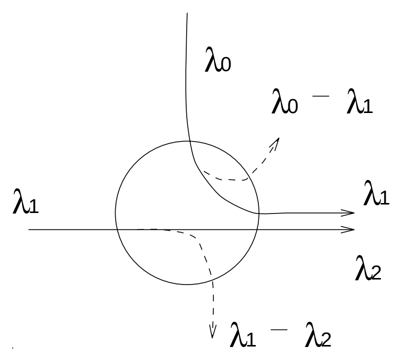

Figure 6: Traffic through a node in the ring

We model the ring of n nodes (Figure 5) by n G/M/N/N queues with correlated arrivals and departures. The queues are numbered $1, 2, \ldots, n$ and are arranged in a ring. A customer arriving to a queue from the outside tries to capture a server. If it fails, it leaves the system, if it succeeds it immediately tries to capture a server in the next queue in the ring. In case of a failure in the second attempt the customer holds the server it captured in the first queue for a period that is exponentially distributed with parameter μ, in case of a success the customer holds a server in both queues for a period that is exponentially distributed with parameter μ. Let λ_1 be the average rate of the bursts that arrive from outside the ring and capture a server, λ_2 be the average rate of bursts that succeed to capture a server in two queues, and denote $p = \lambda_2/(\lambda_1 + \lambda_2)$. p is the probability that a customer in a queue entered the previous queue in the ring. We normalize our results by setting $\mu = 1$. λ_2 is, therefor, our desired node throughput.

Due to symmetry of the ring, the traffic through a node in the ring does not depend on the number of ring nodes (Figure 6). Therefore, the analysis of a ring with any number of nodes greater than two is identical. (A two-node ring is an exception since there is a full correlation there between the number of bursts in the two queues.) We thus continue our analysis by setting $n = 3$.

To describe the three-node system as a Markov chain, each state should be comprised of the number of bursts in each queue (link) and the number of bursts that occupies every combination of two queues. This leads to an $(N+1)^6$ state space. To reduce the number of states, we describe the system only by the number of bursts in each queue, and we assume that with probability p a departure is of a burst that occupies two queues. This enables us to solve a continuous time Markov chain with only $(N+1)^3$ states.

We shall next write expressions for the elements of the transition rate matrix, Q. The transition rates from a state due to arrivals are always equal to the arrival rate λ_0, the transitions are to states where a customer captures a server in two neighboring queues or just in one of them if the next queue is full:

$$\begin{array}{rcll} q_{\langle i,j,k\rangle,\langle i+1,j+1,k\rangle} & = & \lambda_0 & i,j<N \\ q_{\langle i,j,k\rangle,\langle i,j+1,k+1\rangle} & = & \lambda_0 & j,k<N \\ q_{\langle i,j,k\rangle,\langle i+1,j,k+1\rangle} & = & \lambda_0 & i,k<N \\ q_{\langle i,N,k\rangle,\langle i+1,N,k\rangle} & = & \lambda_0 & i<N \\ q_{\langle i,j,N\rangle,\langle i,j+1,N\rangle} & = & \lambda_0 & j<N \\ q_{\langle N,j,k\rangle,\langle N,j,k+1\rangle} & = & \lambda_0 & k<N \end{array} \tag{34}$$

The transition rates due to departures depend on the number of customers in the queues. Since a customer in a queue also occupies a server in a neighboring queue with probability p, we have correlated departures from two neighboring queues with probability p multiplied by the number of customers in the less occupied queue, otherwise we have single departures.

$$\begin{array}{rcl} q_{\langle i,j,k\rangle,\langle i-1,j-1,k\rangle} & = & p\min\{i,j\}\mu \\ q_{\langle i,j,k\rangle,\langle i,j-1,k-1\rangle} & = & p\min\{j,k\}\mu \\ q_{\langle i,j,k\rangle,\langle i-1,j,k-1\rangle} & = & p\min\{k,i\}\mu \\ q_{\langle i,j,k\rangle,\langle i-1,j,k\rangle} & = & (i-p(\min\{i,j\}+\min\{i,k\}))\mu \\ q_{\langle i,j,k\rangle,\langle i,j-1,k\rangle} & = & (j-p(\min\{j,k\}+\min\{j,k\}))\mu \\ q_{\langle i,j,k\rangle,\langle i,j,k-1\rangle} & = & (k-p(\min\{k,j\}+\min\{k,i\}))\mu \end{array} \tag{35}$$

The diagonal elements of Q are the departure rates from the states multiplied by -1:

$$q_{\langle i,j,k\rangle,\langle i,j,k\rangle} = -[((i+j+k)(1-p) - p(\min\{i,j,k\} - \max\{i,j,k\}))\mu + \mathcal{A}(\langle i,j,k\rangle)\lambda_0] \tag{36}$$

where $\mathcal{A}$ is the acceptability index of the state, i.e., the number of queues in this state that are able to accept new customers.

To find the stationary probabilities $\pi_{\langle i,j,k\rangle}$ for $i,j,k \in [0,\ 1,\ \ldots,\ N]$ we solve the matrix equation

$$\vec{\pi}Q = 0 \tag{37}$$

along with the probability conservation relation:

$$\sum_{i,j,k\in[0,\ 1,\ \ldots,\ N]} \pi_{\langle i,j,k\rangle} = 1 \tag{38}$$

The solution of equations 37 and 38 yields the stationary probabilities as a rational functions of p. p can be found using the relations:

$$\lambda_1 = \lambda_0 \sum_{\substack{j,k\in[0,\ 1,\ \ldots,\ N] \\ i\in[0,\ 1,\ \ldots,\ N-1]}} \pi_{\langle i,j,k\rangle} \tag{39}$$

$$\lambda_2 = \lambda_0 \sum_{\substack{k\in[0,\ 1,\ \ldots,\ N] \\ i,j\in[0,\ 1,\ \ldots,\ N-1]}} \pi_{\langle i,j,k\rangle} \tag{40}$$

and the definition of p. This yields the polynomial equation:

$$\sum_{\substack{k\in[0,\ 1,\ \ldots,\ N] \\ i,j\in[0,\ 1,\ \ldots,\ N-1]}} \pi_{\langle i,j,k\rangle} = p\left(\sum_{\substack{j,k\in[0,\ 1,\ \ldots,\ N] \\ i\in[0,\ 1,\ \ldots,\ N-1]}} \pi_{\langle i,j,k\rangle} + \sum_{\substack{k\in[0,\ 1,\ \ldots,\ N] \\ i,j\in[0,\ 1,\ \ldots,\ N-1]}} \pi_{\langle i,j,k\rangle}\right) \tag{41}$$

The value of p can then be substituted in equations 39 and 40 to obtain the values of λ_1 and λ_2. The analysis results were matched by results from simulation of a ring with eight nodes for N=2 and N=4. Figure 7 shows λ_1 and λ_2 as a function of the load $\rho = \lambda_0/(N\mu)$. Note that λ_2 is significantly lower than 0.5, a throughput that can theoretically be achieved for this system.

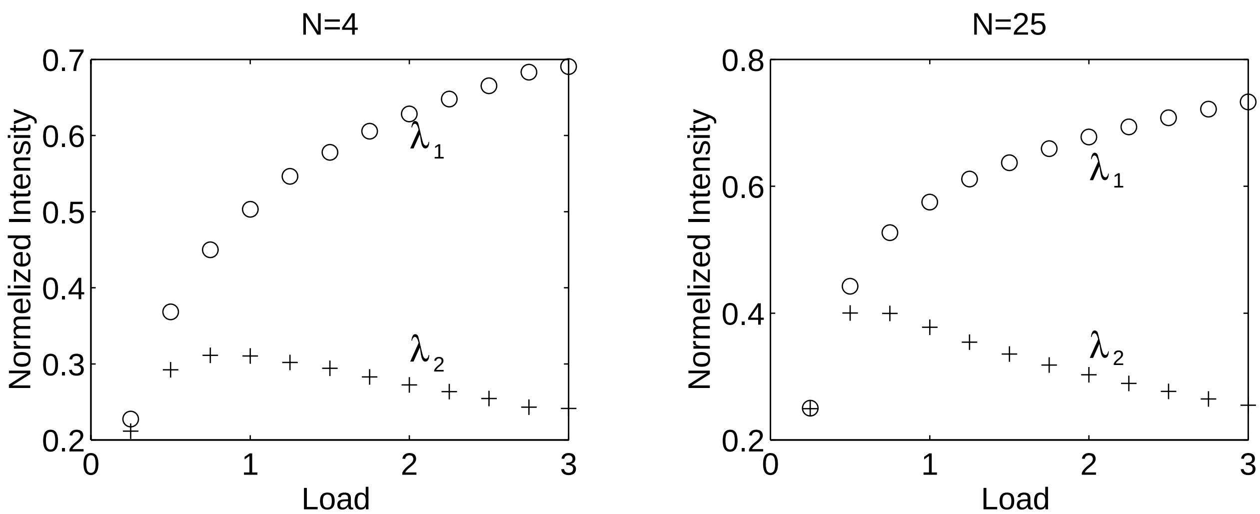

Figure 7: λ_1 and λ_2 as a function of the load, ρ.

Examining the values of the plotted points in Figure 7 yields the relation $\lambda_2/\lambda_1 > \lambda_1/\lambda_0$. This means that bursts that have already captured a server have higher probability of capturing a second server compared to bursts that come from the outside, which contradicts Widjaja's assumption [Wid94] that the probability of capturing a server is constant along the burst route. This probability increases because the traffic carried on a link becomes smoother at every stage. The smoothing is most apparent in the first stage as depicted in Figure 8. The Figure shows the success probability for a burst in successively capturing a server in each of the hops along its route. The points are taken from a simulation of a ring with eight nodes, where each link can support four bursts simultaneously. It is clear that while the success probability increases along the route, the difference is most pronounced between the first and the second hop.

Journal of High Speed Networks, 5(4):347--363, 1996

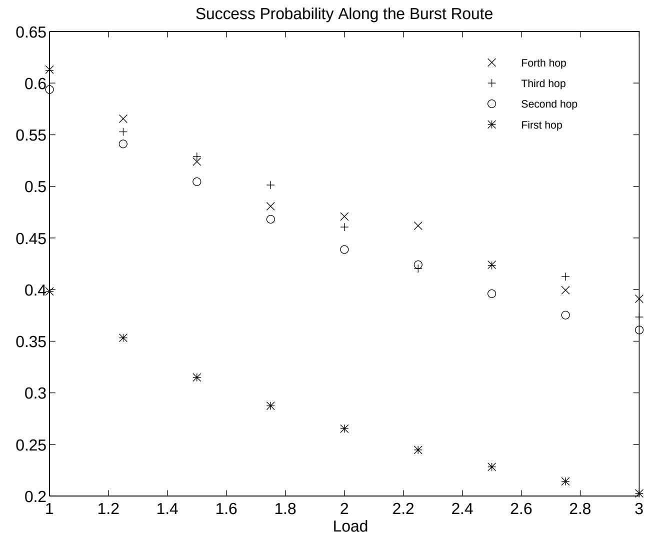

Figure 8: The success probability along the burst route (N=4, h=4).

4 Comparison to Fast Reservation Protocols (FRP)

In this section, we compare the performance of the one-way reservation protocols with the performance of fast reservation protocols [BT92] for the three-node subnetwork and for the ring. As for the one-way-reservation analysis, we assume zero delay on the lines. Bursts in FRPs behave exactly as telephone calls in telephone switching networks, thus, their performance analysis can be done efficiently using results obtained in [**?**, **?**] and the references therein. In appendix A, we give the transition rate matrices we used to obtain our results.

4.1 Comparison for the Three Node Subnetwork

We consider here the system of figure 1. Figure 9 compares the system utilization for the one-way reservation scheme that was computed in section 2.1 (+s) with the utilization for the FRP (os). The solid lines in the graphs show the utilization of an M/M/N/N queue with input rate $\lambda = \rho N \mu$. This is the maximum utilization that can be achieved for input streams 1 and 3 regardless of the capacity of links 3 and 4. Note that as the load increases both reservation schemes increase the link utilization, however when the load approaches infinity the FRP approaches full link utilization, i.e., the value 1, while the one-way reservation protocol is always below the plotted bound (Eq. 32).

4.2 Analysis of FRP in a Cycle of Streams

We consider the system of figures 5 and 6 compare the results obtained in section 3 with the FRP in a ring of size three. The number three is selected since it gives the worst case utilization (see Appendix B). Note that for the one-way-reservation the utilization was the same for all ring sizes.

Figure 10 compares the analysis with the simulation results and depicts the fact that a ring of size 3 has the worst throughput. Figure 11 compares the ring utilization for the one-way reservation scheme that was computed in section 3 ($\times$s) with the ring utilization for the FRP (os). For all

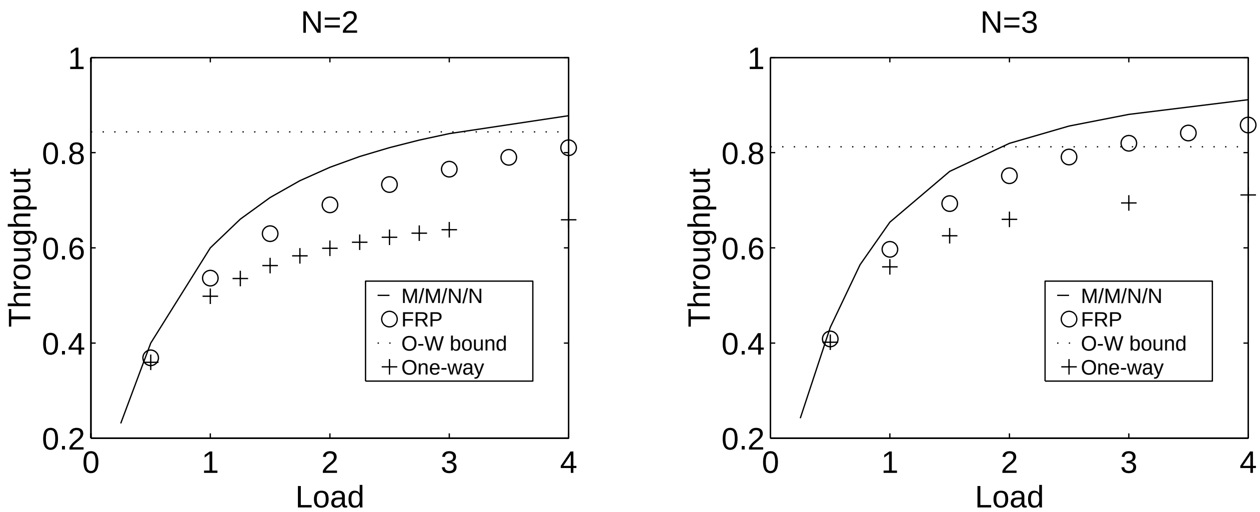

Figure 9: Comparison of analysis results for one-way reservation protocol, FRP, and maximum possible throughput.

the loads, the FRP exhibits better utilization than the one-way reservation. Note that as the load increases the utilization increases for FRP while it decreases for one-way reservation.

5 Discussion and Concluding Remarks

When using one-way reservation for the systems analyzed and for all loads, a large portion of the unsuccessful bursts is rejected at the entry points before any network resources are consumed. This phenomenon is mostly due to the homogeneous systems we analyzed, i.e., systems with identical link capacity. In practice, bursts use the residual capacity left after reservations of constant-bit-rate applications were made, and thus the capacity they see is likely to be different for every link. In non-homogeneous systems we expect to see a larger portions of the bursts that make it through the first hop and subsequently fail.

Altogether, on-the-fly reservation appears not to be an efficient way to reserve bandwidth (see figures 9 and 11) and should be used only when alternatives are impractical. It is clear that it is more suitable for the case where the burst size is small compared to the available throughput or when the system load is low. However, enhancements to this scheme [CRS94] can increase its performance and make it more appealing for use.

The comparison between the two types of reservation protocols shows an advantage for the FRP when the delay on the lines is zero contradicting Cho and Leon-Garcia [CLG94] claims. Future work is required to compare the two reservation schemes when line delay is accurately accommodated into the model.

Journal of High Speed Networks, 5(4):347--363, 1996

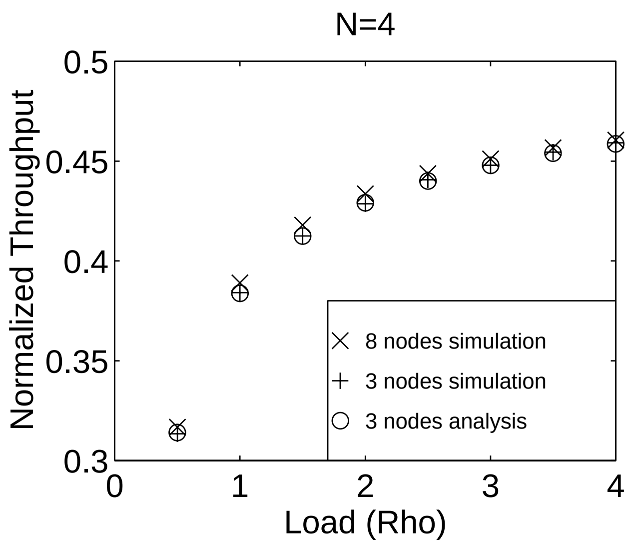

Figure 10: FRP: Comparison of the analysis with simulation results for an eight-node ring.

Appendices

A FRP Analysis

A.1 The Transition Rate Matrix for the Three Node Subnetwork

The elements of the transition rate matrix, Q, are

$$\begin{aligned} q_{\langle i,j,l,m\rangle,\langle i-1,j,l-1,m\rangle} &= l\mu \\ q_{\langle i,j,l,m\rangle,\langle i,j-1,l,m-1\rangle} &= m\mu \\ q_{\langle i,j,l,m\rangle,\langle i-1,j,l,m\rangle} &= (i-l)\mu \\ q_{\langle i,j,l,m\rangle,\langle i,j-1,l,m\rangle} &= (j-m)\mu \end{aligned} \tag{42}$$

For the case where $l+m<N$ and $i+j-(l+m)<N$:

$$\begin{aligned} q_{\langle i,j,l,m\rangle,\langle i+1,j,l+1,m\rangle} &= \lambda/2 \\ q_{\langle i,j,l,m\rangle,\langle i+1,j,l,m\rangle} &= \lambda/2 \\ q_{\langle i,j,l,m\rangle,\langle i,j+1,l,m+1\rangle} &= \lambda/2 \\ q_{\langle i,j,l,m\rangle,\langle i,j+1,l,m\rangle} &= \lambda/2 \end{aligned} \tag{43}$$

For the case where $l+m=N$ and $i+j-(l+m)<N$:

$$\begin{aligned} q_{\langle i,j,l,m\rangle,\langle i+1,j,l,m\rangle} &= \lambda \\ q_{\langle i,j,l,m\rangle,\langle i,j+1,l,m\rangle} &= \lambda \end{aligned} \tag{44}$$

For the case where $l+m<N$ and $i+j-(l+m)=N$:

$$\begin{aligned} q_{\langle i,j,l,m\rangle,\langle i+1,j,l+1,m\rangle} &= \lambda \\ q_{\langle i,j,l,m\rangle,\langle i,j+1,l,m+1\rangle} &= \lambda \end{aligned} \tag{45}$$

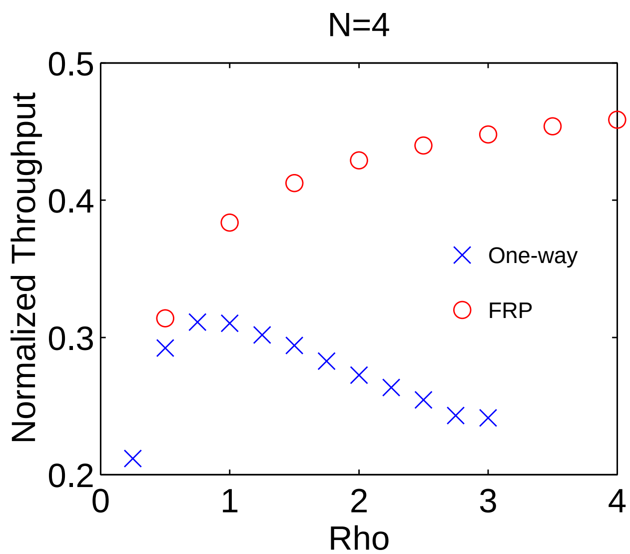

Figure 11: Comparison of the analysis of one-way reservation protocols with FRP for rings.

The elements of the diagonal of Q are

$$q_{\langle i,j,l,m\rangle,\langle i,j,l,m\rangle} = -[2\lambda + (i+j)\mu] \tag{46}$$

For the case where exactly one of the links 3 or 4 is full and exactly one of the links 1 or 2 is full, i.e., ($i = N$ XOR $j = N$) AND ($i + j - (l + m) = N$ XOR $l + m = N$)

$$q_{\langle i,j,l,m\rangle,\langle i,j,l,m\rangle} = -[\lambda/2 + (i+j)\mu] \tag{47}$$

For the case where exactly one of the links 3 or 4 is full and none of the links 1 and 2 is full, i.e., ($i = N$ XOR $j = N$) AND $i + j - (l + m) < N$ AND $l + m < N$

$$q_{\langle i,j,l,m\rangle,\langle i,j,l,m\rangle} = -[\lambda + (i+j)\mu] \tag{48}$$

For the case where exactly one of the links 1 or 2 is full and none of the links 3 and 4 is full, i.e., $i < N$ AND $j = N$ AND ($i + j - (l + m) = N$ XOR $l + m = N$)

$$q_{\langle i,j,l,m\rangle,\langle i,j,l,m\rangle} = -[\lambda + (i+j)\mu] \tag{49}$$

For the case where all the links are full, i.e., $i = j = N$

$$q_{\langle N,N,l,m\rangle,\langle N,N,l,m\rangle} = -2N\mu \tag{50}$$

A.2 The Transition Rate Matrix for the Ring

The system state is described by the number of bursts that enter the system at node i, $i = 1, 2, 3$. The elements of the transition rate matrix, Q, are

$$\begin{aligned}
q_{\langle i,j,k\rangle,\langle i+1,j,k\rangle} &= \lambda_0 \quad i+j<N \quad i+k<N \\
q_{\langle i,j,k\rangle,\langle i,j+1,k\rangle} &= \lambda_0 \quad j+i<N \quad j+k<N \\
q_{\langle i,j,k\rangle,\langle i,j,k+1\rangle} &= \lambda_0 \quad k+i<N \quad k+j<N
\end{aligned}$$

$$
\begin{aligned}
q_{\langle i,j,k\rangle,\langle i-1,j,k\rangle} &= i\mu \\
q_{\langle i,j,k\rangle,\langle i,j-1,k\rangle} &= j\mu \\
q_{\langle i,j,k\rangle,\langle i,j,k-1\rangle} &= k\mu
\end{aligned}
$$

and

$$q_{\langle i,j,k\rangle,\langle i,j,k\rangle} = -[(i+j+k)\mu + \mathcal{A}(\langle i,j,k\rangle)\lambda_0]$$

where $\mathcal{A}$, as in section 3 (Eq. 36), is the acceptability index of the state.

B A proof that the throughput of an n-node ring with FRP, is the lowest when $n = 3$.

We examine unidirectional rings of $n \geq 2$ nodes labeled $0, \ldots, n-1$. Neighboring nodes are connected by links that can support N bursts simultaneously. The links are labeled with the number of the node they emanate from. All bursts travel exactly two hops.

A ring state is associated with a tuple $\langle m_1, m_2, \ldots, m_n\rangle$, where m_i is the number of bursts that enter the ring at node i, and $\bar{m} = \sum_{i=1}^{n} m_i$ is the total number of bursts in the ring. All node indexes are computed modulo n. Note that $\bar{m}/n$ is defined as the normalized ring throughput. We define a *saturation state* to be a state where no new bursts can reserve bandwidth along their entire path before an existing burst releases bandwidth.

Theorem B.1 *The maximum number of bursts that a ring can simultaneously support is upper bounded by* $\lfloor \frac{nN}{2} \rfloor$.

Proof: Each of the n links can support up to N bursts simultaneously. The ring can, thus, support nN single-link bursts. To get the upper bound, this number is divided by the number of links each burst occupies. □

A saturation state that achieves the upper bound of Theorem B.1, i.e., $\bar{m} = \lfloor \frac{nN}{2} \rfloor$, is called *optimal*. An optimal state exist for any ring of size $n \geq 2$ if N is even and exactly $N/2$ bursts enter the system at each node. This state is associated with the tuple $\langle N/2, N/2, \ldots, N/2\rangle$. For simplicity, we assume from now on that N is even.

Lemma B.1
A. For all states, $m_i + m_{i-1} \leq N$.
B. For any saturation states and for all i*, either* $m_i + m_{i-1} = N$ *or* $m_i + m_{i+1} = N$.

Proof: $m_i + m_{i-1}$ is the number of bursts that are carried by link i. Obviously, this number can never exceed N, which proves part A.

To prove part B, suppose that in a saturation state both $m_i + m_{i-1} < N$ and $m_i + m_{i+1} < N$. This means, that both links i and $i+1$ carry less than N bursts, and therefore at least one additional burst can be accepted at node i, contradicting the assumption that this is a saturation state. □

Based on Lemma B.1, the following observations can be made:

1. For rings of size 2 and 4, all the saturation states are of the form $\langle N/2 - i, N/2 + i\rangle$ and $\langle N/2 - i, N/2 + i, N/2 - i, N/2 + i\rangle$, $-N/2 \leq i \leq N/2$, respectively. All these states are optimal.

Journal of High Speed Networks, 5(4):347--363, 1996

2. For a ring of size 3, all the saturation states are of the form $\langle N/2 - i, N/2 + i, N/2 - i \rangle$, $0 \leq i \leq N/2$ (and the cyclic rotation of these states). Note that only for $i = 0$ the state is optimal.

We adhere to the model we use in the analysis in sections 3 and 4.2, i.e., burst duration is exponentially distributed with parameter μ and burst arrivals to each node form a Poisson process (with parameter λ). As in sections 3 and 4.2, the ring can be modeled by a Markov chain whose states are described by the n-tuple $\langle m_1, m_2, \ldots, m_n \rangle$ as defined above. Since this Markov chain is finite with exponential transitions, it is ergodic, meaning that we can replace the average ring utility in time with the utility probability at every moment.

Our goal is to show that rings of size 3 have the lowest throughput. For this aim, we show that, for heavy traffic, the Markov chain of a large ring is a Cartesian product of Markov chains of smaller rings. In our proof, we define heavy traffic as the rate when every burst that releases bandwidth does it from a saturation state. We define a *segment* to be a maximal length sequence of neighboring nodes s.t. for every two neighboring nodes in the sequence, i and $i+1$, $m_i + m_{i+1} = N$.

Lemma B.2 *Every saturation state is comprised of one or more segments of length 2 or more.*

Proof: Stems directly from Lemma B.1. □

Lemma B.3 *Consider a ring in a saturation state that is comprised of K segments $\langle s_1, \ldots, s_i, \ldots, s_K \rangle$, and suppose bandwidth is released by a burst that enters the ring from a node in segment s_i. The transition probability to the next saturation state of the ring is independent of $s_j \neq s_i$, under the heavy traffic assumption.*

Proof: Assume that nodes $i-1$ and i belong to different segments. A departure of a burst that enters the system from node i can not enable the acceptance of new bursts in node $i-1$ since by the segment definition $m_{i-1} + m_{i-2} = N$. Similarly, a departure of a burst that enters the system from node $i-1$ can not enable the acceptance of new bursts in node i. Obviously, a departure of a burst that enters the system from a node that is not on a segment border can not enable the acceptance of new bursts in nodes out of this segment. As a result, the transition probability among saturation states in a segment is independent of other segments and the Lemma holds. □

Theorem B.2 *For heavy traffic, a ring of size 3 has the lowest throughput.*

Proof: Consider a ring of size $k \geq 5$ which is in a saturation state that is comprised of $K \geq 1$ segments, $s_1, \ldots, s_K$. Three types of segments are possible:

1. Even size segments.
2. Odd size segments of the form $\langle N/2 + i, N/2 - i, \ldots, N/2 + i \rangle$, $0 \leq i \leq N/2$.
3. Odd size segments of the form $\langle N/2 - i, N/2 + i, \ldots, N/2 - i \rangle$, $0 \leq i \leq N/2$.

By lemma B.3, the total number of bursts is given by $\sum_{j=1}^{K} m_{s_j} |s_j|$, where $|s_j|$ is the size of segment s_j and $m_{s_j} = \frac{1}{|s_j|} \sum_{k \in s_j} m_k$ is the average number of bursts per node in segment s_j.

For segments of type 1, $m_{s_j} = N/2$. For segments of type 2, $m_{s_j} \geq N/2$. For segments of type 3, the average number of bursts is at least equal to that of a ring of size 3. Thus, the average number of bursts is a weighted sum of three values each of which is equal or greater than the average number of bursts in a ring of three nodes. This holds for every starting saturation state. $\square$

We proved our claim for heavy traffic. For light traffic, almost all arriving bursts succeed to reserve bandwidth regardless of the ring size. Simulations show that our claim holds also for medium loaded rings.

References

[BT92] Pierre E. Boyer and Didier P. Trachier. A reservation principle with applications to the ATM traffic control. *Computer Networks and ISDN Systems*, 24:321 – 334, 1992.

[CLG94] You-Ze Cho and Alberto Leon-Garcia. Performance of burst-level bandwidth reservation in ATM LANs. In *INFOCOM'94*, pages 812 – 820, June 1994.

[CRS94] Israel Cidon, Raphael Rom, and Yuval Shavitt. Fast bypass algorithms for high-speed networks. Technical Report EE PUB No. 924, Technion - Israel Institute of Technology, June 1994. Submitted for publication.

[Gir90] André Girard. *Routing and Dimensioning in Circuit-Switched Networks*. Addison-Wesley, 1990.

[HLNP93] Daryoush Habibi, D. J. H. Lewis, D. T. Nguyen, and J. Pieloor. Analysis of an access node multiplexer in a system serving CBR and VBR traffic. *Computer Communications*, 16(12):776 – 780, December 1993.

[Kle75] Leonard Kleinrock. *Queueing Theory*, volume I. Wiley, 1975.

[Kle76] Leonard Kleinrock. *Queueing Theory*, volume II. Wiley, 1976.

[Tak62] Lajos Takács. *Introduction to the Theory of Queues*. Oxford University Press, 1962.

[Tur92] Jonathan S. Turner. Managing bandwidth in ATM networks with burtsy traffic. *IEEE Network*, 6(5):50 – 58, September 1992.

[Wid94] I. Widjaja. Random access for ATM LANs and WANs. In *ICC'94*, May 1994.